

GSA TRAVEL DATA CHALLENGE

gsa.greggparrish.com

github.com/greggparrish/gsa

GREGG PARRISH

grp224@nyu.edu

ABOUT THE SITE

I've organized the site to accommodate three primary uses: as an analytical tool to track expenses; a comparative guide to suggest less expensive regional options; and an example of a portal which could be developed to connect government travelers with the resources and information they need to find inexpensive and reliable travel arrangements.

THE HOMEPAGE

I wanted to keep the initial page fairly simple and let the user choose the focus rather than distract them with too many options and unnecessary information. The only exception being the information along the bottom which are problem areas where a lot of money is being spent, or high traffic routes where greater attention could provide the most potential return.

FLIGHT PAGES

The data is presented by searching for a specific flight route. With little effort, the site could be expanded to include overview pages for individual airports (not JFK >> LAX, but simply JFK), with local expenses and other options being provided, but I didn't see this as being the most immediately useful at first.

The chart at the top tracks ticket prices so you can see any patterns in seasonal pricing.

The columns below that juxtapose past flights with potential future flights. Existing data is presented first as text with clear red and green markers to give an instant impression of where money is spent and saved. These were written to be faster for users to understand than tracking data in a table, although that is also provided lower on the page.

All of the text changes automatically based on the results of the search and will continue to offer suggestions and comparisons as new data is entered into the database.

SUMMARY OF PAST FLIGHTS

AVERAGE TICKET PRICE

National Average (DoT)	Government Paid (GSA)
\$498	\$274

OVERVIEW

Overall, government agencies spent 45% (\$224 per ticket) **less than** the national average on flights from John F Kennedy Intl (New York) to Los Angeles Intl (Los Angeles).

NEARBY AIRPORTS

This route is usually not the least expensive out of all of the combinations of area airports. If possible, it might be cheaper to fly from LGA to BUR. For a complete list of nearby airports and their average fares (DoT), see below.

CITY PAIRS

13 flights on this route (61% of total) took advantage of the City Pairs Program. On average, these fares were \$281, which was **\$7 more expensive** per ticket than non-City Pairs flights, and **\$217 less than** the national average.

TICKETS BOUGHT ONLINE

3 tickets were purchased online (14% of total), and with an average price of \$208, were **\$66 less expensive** per ticket than those that were bought through other means.

RESOURCES FOR FUTURE FLIGHTS

CITY PAIRS PROGRAM

There is a **discount fare** available for this route on United through the City Pairs Program.

Unrestricted coach fare: \$167.

Capacity controlled fare: \$111.

Use the CPP's search tool for details and more rates.

PLANNING A CONFERENCE?

Recently, flights from JFK to Cody, WY (\$105.66), Baltimore, MD (\$110.29) and Milwaukee, WI (\$112.46) have been the least expensive.

Inexpensive flights to LAX are often from Imperial, CA (\$75.56), Carlsbad, CA (\$86.69) and Yuma, AZ (\$87.69).

HOTELS AND RENTAL CARS

See data for local expenses around New York or Los Angeles.

CURRENTLY AVAILABLE FARES

City Pair (O/W)	Market (R/T)
\$111	\$318*

*This fare is from Kayak.com, provided as proof of concept and is in accordance with their ToS. It is accessed through a simple RSS call and could easily be switched out for another service, or made more accurate by asking Kayak for direct API access.

NEARBY AIRPORTS

A feature that I think is particularly helpful is the search of nearby airports. The site will search all of the airports within 30 miles of both your origin and destination and offer the cheapest combination.

NEARBY AIRPORTS

This route is usually not the least expensive out of all of the combinations of area airports. If possible, it might be cheaper to fly from **LGA to BUR**. For a complete list of nearby airports and their average fares (DoT), see below.

In the statistics box below, these flight combinations are ordered according to price and listed for the user to compare. Clicking on any of the routes will take you to the page for that flight.

THIS ROUTE COMPARED TO AIRPORTS WITHIN 30 MILES

LGA - BUR	LGA - LGB	HPN - LAX	JFK - LGB	EWB - LGB
\$241.06	\$256.70	\$267.46	\$268.66	\$270.99

STATISTICS

The statistics section organizes the most commonly needed data – totals, averages, high and low outliers – into a few simple tables, but gives the option to see all of the data at the bottom.

LOCAL EXPENSES

The maps are drawn by finding the closest FedRooms hotels to any given city center. Clicking on any of them will reveal the telephone number for reservations. The list on the right is randomized to offer additional options for travelers and make it easier to choose one of these less expensive options that finding their own hotel online.

Similar to the flight routes, these local pages search past hotel reservation data in nearby cities to compare the average regional price. This allows for the option of staying outside of the city center where it is typically more expensive, especially if the work being addressed is in one of the surrounding towns.

ADMIN INTERFACE

Django comes with an admin panel making it easy to add new records or change existing ones.

Change tmis_and_smartpay_flights

History

Date	2010-08-25	
Fare	489.40	
Airline	Southwest	
Origin	MCI	+
Dest	SAN	+
Online	N	
Citypair	Economy Restricted Fare (Non-Contract)	

Delete

Save and continue editing

Save and add another

Save

FURTHER RECOMMENDATIONS

If the spreadsheets provided by the GSA for this challenge are representative of the state of existing GSA travel data, then the most help that programmers could offer is in data collection. There are a few fields that could be added to assist in tracking users and identifying how more money could be saved, but if that data is collected as these were, then they will just add to the problem.

The challenge asked us to make a tool that would be easy to update, even for those who don't have any programming experience. The site I created allows for easy import of CSV files, but uploading these files would then render the data tools useless. It would for any site. Cities are spelled any number of ways, hotel names might exist or could be addresses or phone numbers, many fields aren't filled out, some people checked out of hotels before checking in, or dropped off cars before they took them, and a few apparently rented cars for over a year for just pennies a day. Almost all of these are simple mistakes, some might be due to deals arranged by a travel agent or through some sort of program, but together they make it difficult for an automated process to aggregate the data into any sort of meaningful results.

This must cause numerous problems for the GSA and a lot of human effort to either clean the data or limit the ability of bad data to alter the results.

Ideally, I think the GSA could use a site for data analysis that also worked as a portal for flight, hotel and car reservations. This isn't anywhere near as difficult as it sounds and would not require the GSA to handle the actual transactions, but it would encourage more people to consider the money being spent and choose from a list of less expensive options.

Most large search providers, like Priceline, Kayak, Orbitz and Expedia, all offer APIs that would allow the GSA to create a site for users to book FedRooms, or use a meta-search engine like Kayak to find the absolute cheapest available flight.

Placing all of these options next to a comparative summary of how much *should* be spent on that flight will instantly let people know if they are spending too much.

From the perspective of data analysis, a reservation portal could process most of the data, so city names would be standardized, dates and totals correct, and the proper fields entered all without bothering the user to fill out another series of forms.

From a technical angle, a portal like this wouldn't be a difficult project, but I assume there are a large number of things that might prevent this on the government side. Given that, the forms used to collect data could also be rewritten to fill out as much of a form as possible with autocompletes, dropdowns, required fields or other methods of standardization. I don't know how these are currently done so it is difficult to make recommendations, but generally they should rely as little as possible on user input.

TECHNICAL DETAILS

The site is written in Python using Django as a framework, MySQL for the database, NVD3 for charts, Google Maps on the local pages, and JQuery for the autocomplete on the search and the blue navigation bars on the interior pages. Except for the Google maps, the site has all been written using open source technology and is incredibly flexible.

The main body of code that crunches through the math could be pulled into any Python project with just a little tweaking, or could be ported fairly easily into another language and framework.

Despite the considerable amount of data being processed and the relatively large number of overall records (the DoT averages number almost a million), I've avoided bottlenecks by limiting database access, putting resource intensive processes into maintenance scripts so the data is there when needed, and letting Python do the math rather than the database. The current host is just a low tier EC2 but is still quite fast. On a normal server, speed would not be an issue at all.

I've released it under a GPL v3 license and it is available on Github: github.com/greggparrish/gsa